Diffusion Self-Guidance for Controllable Image Generation

Abstract

Large-scale generative models are capable of producing high-quality images from detailed text descriptions. However, many aspects of an image are difficult or impossible to convey through text. We introduce self-guidance, a method that provides greater control over generated images by guiding the internal representations of diffusion models. We demonstrate that properties such as the shape, location, and appearance of objects can be extracted from these representations and used to steer sampling. Self-guidance works similarly to classifier guidance, but uses signals present in the pretrained model itself, requiring no additional models or training. We show how a simple set of properties can be composed to perform challenging image manipulations, such as modifying the position or size of objects, merging the appearance of objects in one image with the layout of another, composing objects from many images into one, and more. We also show that self-guidance can be used to edit real images. For results and an interactive demo, see our project page at https://dave.ml/selfguidance/